What is claimed is:

1. A method to identify topics in a data corpus having a plurality of segments, comprising:

determining a segment-level actual usage value for one or more word combinations;

computing a segment-level expected usage value for each of the one or more word combinations; and

designating a word combination as a topic if the segment-level actual usage value of the word combination is substantially greater than the segment-level expected usage value of the word combination.

2. The method of claim 1, wherein each of the plurality of segments comprises a portion of a document.

3. The method of claim 2, wherein the portion of a document comprises a paragraph.

4. The method of claim 2, wherein the portion of a document comprises a heading.

5. The method of claim 2, wherein the portion of a document comprises the entire document.

6. The method of claim 1, wherein each of the one or more word combinations comprise two or more substantially contiguous words.

7. The method of claim 6, wherein two words are substantially contiguous if they are separated only by zero or more words selected from a predetermined list of words.

8. The method of claim 7, wherein the predetermined list of words comprise STOP words.

9. The method of claim 1, wherein at least one word in each of the one or more word combinations is selected from a predetermined list of words.

10. The method of claim 9, wherein the predetermined list of words comprise a list of domain specific words.

11. The method of claim 1, wherein the act of determining a segment-level actual usage value for a word combination comprises determining the number of segments in the data corpus the word combination is in.

12. The method of claim 1, wherein the act of computing a segment-level expected usage value for each of the one or more word combinations comprises calculating a value in accordance with:

$$\frac{S(w_i)\times S(w_j)\times\ldots\times S(w_m)}{N^{m-1}},$$

where "m" represents the number of words in the word combination, "N" represents the number of segments in the data corpus, and $S(w_i)$ represents the number of unique segments in the data corpus that word w_i of the word combination is in.

13. The method of claim 1, wherein the act of designating a word combination as a topic, comprises designating a word combination as a topic if the segment-level actual usage value of the word combination is greater than approximately twice the segment-level expected usage value of the word combination.

14. The method of claim 1, wherein the act of designating a word combination as a topic, comprises designating a word combination as a topic if the segment-level actual usage value of the word combination is greater than a specified value.

15. The method of claim 14, wherein the act of designating a word combination as a topic, comprises designating a word combination as a topic if the segment-level actual usage value of the word combination is greater than approximately 10.

16. A program storage device, readable by a programmable control device, comprising instructions stored on the program storage device for causing the programmable control device to identify topics in a data corpus having a plurality of segments, the instructions causing the programmable control device to:

determine a segment-level actual usage value for one or more word combinations;

compute a segment-level expected usage value for each of the one or more word combinations; and

designate a word combination as a topic if the segment-level actual usage value of the word combination is substantially greater than the segment-level expected usage value of the word combination.

17. The program storage device of claim 16, wherein the instructions for identifying topics in segments comprise instructions to identify topics in a portion of a document.

18. The program storage device of claim 17, wherein the instructions to identify topics in a portion of a document comprise instructions to identify topics in a paragraph.

19. The program storage device of claim 17, wherein the instructions to identify topics in a portion of a document comprise instructions to identify topics in an entire document.

20. The program storage device of claim 16, wherein the instructions to designate a word combination as a topic comprise instructions to designate word combinations of two or more substantially contiguous words.

21. The program storage device of claim 20, wherein the instructions to designate two or more substantially contiguous words as a topic comprise instructions to designate two or more words if they are separated only by zero or more words selected from a predetermined list of words.

22. The program storage device of claim 16, wherein the instructions to designate a word combination as a topic comprise instructions to designate a word combination as a topic only if at least one of the designated words is selected from a predetermined list of words.

23. The program storage device of claim 22, wherein the instructions to designate words from a predetermined list of words comprise instructions to select words from a domain specific word list.

24. The program storage device of claim 16, wherein the instructions to determine a segment-level actual usage value for a word combination comprise instructions to determine the number of segments in the data corpus the word combination is in.

25. The program storage device of claim 16, wherein the instructions to compute a segment-level expected usage value for each of the one or more word combinations comprise instructions to calculate a value in accordance with:

$$\frac{S(w_i)\times S(w_j)\times\ldots\times S(w_m)}{N^{m-1}},$$

where "m" represents the number of words in the word combination, "N" represents the number of segments in the data corpus, and $S(w_i)$ represents the number of unique segments in the data corpus that word w_i of the word combination is in.

26. The program storage device of claim 16, wherein the instructions to designate a word combination as a topic, comprise instructions to designate a word combination as a topic if the segment-level actual usage value of the word combination is greater than approximately twice the segment-level expected usage value of the word combination.

27. The program storage device of claim 16, wherein the instructions to designate a word combination as a topic, comprise instructions to designate a word combination as a topic if the segment-level actual usage value of the word combination is greater than a specified value.

28. The program storage device of claim 27, wherein the instructions to designate a word combination as a topic, comprise instructions to designate a word combination as a topic if the segment-level actual usage value of the word combination is greater than approximately 10.

29. A method to display a list of topics associated with data items stored in a database, comprising:

identifying a result set based on an initial user query, the result set identifying a plurality of stored data items;

identifying those topics associated with the stored data items identified in the result set;

selecting for display a topic associated with the most identified stored data items;

selecting for display another topic, said another topic associated with the most identified stored data items not associated with a previously identified display topic, wherein this step is repeated until all identified stored items in the result set have been accounted for; and

displaying the selected display topics.

30. The method of claim 29, wherein the act of identifying a result set comprises:

identifying an initial result set, the initial result set identifying a first plurality of stored data items; and

selectively identifying a subset of the initial result set as the result set.

31. The method of claim 30, wherein the act of selectively identifying comprises randomly selecting a specified portion of the initial result set.

32. The method of claim 31, wherein the act of randomly selecting comprises randomly selecting approximately one-percent of the initial result set.

33. The method of claim 29, wherein the act of identifying those topics associated with the stored data items identified in the result set, comprises generating a list of unique topics associated with the identified stored data items.

34. The method of claim 33, further comprising, removing from the generated list those topics that are associated with more than a specified fraction of the identified stored data items.

35. The method of claim 34, wherein the act of removing comprises removing from the generated list those topics that are associated with more than approximately eighty-percent (80%) of the identified stored data items.

36. The method of claim 29, further comprising, displaying a selected number of stored data item identifiers.

37. The method of claim 36, wherein the act of displaying a selected number of stored data item identifiers, comprises displaying a hyperlink.

38. The method of claim 29, wherein the act of selecting for display another topic, comprises determining when the number of data items not associated with a previously identified display topic is less than a specified value and, when this is true:

generating a list of unique individual words from the topics not yet selected for display,

selecting for display a unique word from the list of unique individual words associated with the most identified stored data items; and

selecting for display another unique word from the list of unique individual words, said another unique word associated with the most identified stored data items not associated with a previously identified display topic and unique word, wherein this step is repeated until all identified stored items in the result set have been accounted for.

39. A program storage device, readable by a programmable control device, comprising instructions stored on the program storage device for causing the programmable control device to display a list of topics associated with data items stored in a database, the instructions causing the programmable control device to:

identify a result set based on an initial user query, the result set identifying a plurality of stored data items;

identify those topics associated with the stored data items identified in the result set;

select for display a topic associated with the most identified stored data items;

select for display another topic, said another topic associated with the most identified stored data items not associated with a previously identified display topic, wherein this step is repeated until all identified stored items in the result set have been accounted for; and

display the selected display topics.

40. The program storage device of claim 39, wherein the instructions to identify a result set comprise instructions to:

identify an initial result set, the initial result set identifying a first plurality of stored data items; and

selectively identify a subset of the initial result set as the result set.

41. The program storage device of claim 40, wherein the instructions to selectively identify comprise instructions to randomly select a specified portion of the initial result set.

42. The program storage device of claim 41, wherein the instructions to randomly select comprise instructions to randomly select approximately one-percent of the initial result set.

43. The program storage device of claim 39, wherein the instructions to identify those topics associated with the stored data items identified in the result set, comprise instructions to generate a list of unique topics associated with the identified stored data items.

44. The program storage device of claim 43, further comprising instructions to remove from the generated list those topics that are associated with more than a specified fraction of the identified stored data items.

45. The program storage device of claim 44, wherein the instructions to remove comprise instructions to remove from the generated list those topics that are associated with more than approximately eighty-percent (80%) of the identified stored data items.

46. The program storage device of claim 39, further comprising instructions to display a selected number of stored data item identifiers.

47. The program storage device of claim 46, wherein the instructions to display a selected number of stored data item identifiers, comprise instructions to display a hyperlink.

48. The program storage device of claim 39, wherein the instructions to select for display another topic, comprise instructions to determine when the number of data items not associated with a previously identified display topic is less than a specified value and, when this is true:

generate a list of unique individual words from the topics not yet selected for display,

select for display a unique word from the list of unique individual words associated with the most identified stored data items; and

select for display another unique word from the list of unique individual words, said another unique word associated with the most identified stored data items not associated with a previously identified display topic and unique word, wherein these instructions are repeated until all identified stored items in the result set have been accounted for.